The Noetic Prism: A New Perspective on the Information, Data, Knowledge Complex

Diarmuid Pigott and Valerie Hobbs

School of Information Technology
Murdoch University
E-mail: {d.pigott, vhobbs}@murdoch.edu.au

ABSTRACT

Successful knowledge management has as prerequisite an understanding of what actually constitutes 'knowledge'. Yet the definitions of knowledge in the literature, and its relationships with 'data' and 'information', are varied, inconsistent and often contradictory. In particular the traditional hierarchy of data-information-knowledge and its various revisions do not stand up to close scrutiny. We suggest that the problem lies in a flawed analysis that sees data, information and knowledge as separable concepts that are transformed into one another through business processing. We propose the term 'noetica' to describe the collective intellectual resources of an organisation, together with a revised analysis of the process of adding value to noetica that identifies it as change occurring simultaneously along three different dimensions: increase in aggregation; increase in shape, and increase in contextualisation. A new model is presented in which the three dimensions of complexity – granularity, shape and scope – are seen as the three vertices of a triangular prism, and suggest that all value-adding through business processing can be seen as movement within this space. We map these definitions back to the concepts of information, data and knowledge, and show how the 'noetic prism' can be used as a management tool for analysing the state of the noetic resources of an organisation, and for elucidating and suggesting solutions to several of the current problems in information/data/knowledge management.

Keywords: *knowledge management, knowledge hierarchy, complexity, noetica, noetic prism*

INTRODUCTION

Knowledge management is indisputably one of the hot topics in IT today. However, there seems to be no agreement as to what exactly defines this 'knowledge' that we are managing. In their recent review, Alavi and Leidner (2001) describe the many suggested taxonomies of knowledge, which range from an 'object' at the end of a process model of computation, to an ineffable social construct known only through personalisation. Where consensus exists, it is that the management of knowledge resources is of critical importance in a resource-scarce economy where the efficacy of a digital background is seen as the chief indicator of success or failure of an enterprise. Alavi & Leidner conclude that there is no single approach to knowledge management possible, and suggest that a variety of approaches and systems needs to be employed in an organisation to deal with the diversity of knowledge types and characteristics.

It has been argued that an understanding of the relationship between knowledge and its traditional relatives data and information is essential in knowledge management, and that a lack of understanding can lead to problems in information systems design and costly mistakes in creating inappropriate solutions (Davenport & Prusak, 1998; Sveiby, 1997). The traditional hierarchy of data-information-knowledge has come under scrutiny in recent years, but there is still no agreed definition of the roles or relationships of these three principle components of the IT resource.

In this paper we re-examine the conventional definitions of data, information and knowledge, not only as they are commonly used today, but also as they first arose in the literature. We offer a revised analysis of the nature of the intellectual resources of an organisation and of the process of value-adding to them that is central to business processing. We then present a new model, the *noetic prism*, to describe the dimensions of complexification of these intellectual resources. Finally, we show how the noetic prism can be applied as a management tool to the analysis and solution of many of the pressing resource management issues facing organisations today.

DATA, INFORMATION, KNOWLEDGE

The traditional distinction between data and information is met with in nearly all introductory texts on IT: data is seen as raw material, and information as data that has been processed in some way to become useful. A typical definition is that provided by Hutchinson and Sawyer (2000): "…data consists of the raw facts and figures that are processed into information. Information is summarised data or otherwise manipulated (processed) data." This notion of data as the raw stuff of computation is also the one traditionally met with in the experimental sciences, where it is the basic material for collection and analysis. Database texts (e.g. Elmasri & Navathe, 1999) generally extend the definition of data to encompass its formal structuring within a database, and implicitly acknowledge the intellectual effort involved in this, but their definition still has data as the 'raw' material that is not in itself useful.

An early definition of information arises from the work of Weiner (1948) and the early formulation of cybernetics: Shannon and Weaver (Shannon, 1948; Weaver & Shannon, 1949) define it in terms of communication theory, as a measure of uncertainty in a message stream, and explicitly not its meaning. Weaver and Shannon (1949) redefined it in the same work as "a measure of one's freedom of choice in selecting a message", but the inherency of meaning is still missing from it.

The definition of information as relating to meaning is first formulated by Mooers, when he coined the term 'information retrieval': "When we speak of information retrieval, we are really thinking about the use of machines in information retrieval. The purpose of using machines here, as in other valid applications, is to give the machines some of the tasks connected with recorded information that are most burdensome and unsuited to performance by human beings" (Mooers, 1950). Mooers was an alumnus of Bush at MIT, and it is Mooers' restatement of Bush's idea of information as something that assists the power of thought (Bush, 1945) which is the definition most often met with in current information systems usage.

The functionality of *databases* as we now know them emerged slowly in the late 1950s, from COMPOOL at MIT in 1955, to the IBM Formatted File System (FFS) in 1961, which is the first system with a functional, persistent schema (North, 1999), and by 1957, the US Department of Defence had formed the Conference on *Data* Systems Languages (CODASYL) to standardise industry practice. However, the use of terms involving data (data, data retrieval, data store, data systems, data banks), and their relation to those involving information (information retrieval, information stores, information systems) is by no means clear: simultaneously the terms are used for high- and low-level systematics by different groups, as we can see from the literature. The first commercially released DBMS was IBM's Generalized *Information* Retrieval and Listing System (GIRLS) in 1962, while the 1963 add-on functionality to COBOL was called the *Information Data* Store (IDS). In 1969 Sammet, in discussing the functional sections of programs defines data as "the elements on which the computation is to be performed" (Sammet, 1969), but Codd's landmark paper of relational theory refers to the operation of universals on "data banks" (Codd, 1970).

In the current literature, information is generally seen as something that has been processed or contextualised and which can form the basis for decision or action: "Data becomes information when its creator adds meaning" (Davenport & Prusak, 1998); "Information is data that has value. Informational value depends on context. Until it is placed in an appropriate context, data is not information, and once it ceases to be in that context it ceases to be information" (Clarke, 1999). Date (1991) recognises a similar difference between information and data, but believes the terms themselves are essentially synonymous: "Some writers prefer to distinguish between the two, using 'data' to refer to the values actually stored in the database and 'information' to refer to the meaning of those values as understood by some user. The distinction is clearly important – so important that it seems preferable to make it explicit, where relevant, instead of relying on a somewhat arbitrary differentiation between two essentially similar terms."

The notion that the distinction between data and information depends on its immediate context of use is also frequently stated. Hutchinson & Sawyer (2000) go on to say: "…one person's information may be another person's data. The information of paychecks and payrolls may become data that goes into someone's yearly financial projections or tax returns", and of course this is the basis for the traditional levels of management information systems that supply ever more refined data/information, from the transaction processing system capturing daily business operations to the highly summarised and contextualised executive information system used for long term strategic decision making.

It is much harder to isolate the emergence of the term 'knowledge' into the literature. Clearly it comes from both management science and artificial intelligence research, and can be first met with conceptually in Simon's *Science of the Artificial* (Simon, 1969; 1995). By the early 1980s, the problem of coping with information (foreseen by Mooers 20 years earlier) has led to a specialised literature (e.g. Masuda, 1981), and it is with Newell's *The Knowledge Level* (Newell, 1982) that the formal separation of computable knowledge-tasks is complete.

Knowledge generally enters the traditional models at a higher plane than data/information. In its simplistic definition it is the 'next' level up from data and information, with the value-adding this time achieved by human reasoning or judgement. Hutchinson & Sawyer again: "…knowledge is the result of reasoned analysis of information – a set of organised statements of facts or ideas, communicated in some systematic form". Davenport & Prusak (1998) state that "knowledge derives from information as information derives from data", and see personal values and beliefs as integral to knowledge. Other definitions stress the human and situated aspects of knowledge. Polanyi (1966) considered that all knowledge has a tacit and explicit component, with knowing emerging as a dynamic interaction between focal and unarticulated components of meaning. Clancey (1997), in a wide ranging reappraisal of the dominant paradigm in knowledge representation for the last 20 years (and one in which he participated), argues for knowledge as irreducibly based in human semantic spaces that can not be considered simply a property derivable from disembodied associations: this requirement for an embodiment of knowledge has yet to make it into the business computing literature.

It should be noted that some authors (Bellinger, 1997; Pór, 2000) extend the model to include wisdom: "…utilizing new ways to channel raw data into meaningful information. That information, in turn, can then become the knowledge that leads to wisdom" (Alberthal, 1995; quoted in Bellinger). It would seem that these uses for *wisdom* are required by a somewhat restrictive usage of knowledge, as explicitly removing the possibility of higher order analysis from the domain of knowledge. This practice is, however, not widespread. Another extension met with in the literature is "intelligence" derived from the older term for the composite business of military, governmental and business espionage. Recast by the Gartner group in the late 1980s it was used as a coverall term for "the class of applications and tools by which end-users without a high degree of computer literacy may access, analyse and act upon information" (quoted in Newing, 2000). Again, this is a division required by definition and is not sufficiently used to be significant.

Several authors have disputed the one-way relationship from data to knowledge, pointing out that knowledge work is necessary for the selection of data and the processing involved in its conversion to information. Roszak's (1986) "No ideas, no information" and Miller's (1999) assertion that "information has no intrinsic meaning" indicate a primary role of knowledge in all action. Alavi & Leidner (2001) criticise the traditional hierarchy and state that the key to differentiating between knowledge and information is that "information is converted to knowledge once it is processed in the mind of individuals and knowledge becomes information once it is articulated and presented in the form of text, graphics, words or other symbolic forms". Tuomi (1999) reverses the traditional hierarchy completely, with knowledge as a prerequisite for information, and information for data. For Tuomi, knowledge must be articulated, verbalised and structured to become communicable information, and data is created from information by putting it into a structured form that can be automatically processed.

Alternative hierarchical models also exist. Earl (1994) describes four levels, each level representing an increasing amount of structure, certainty and validation: events are collected and processed to become data, which is further manipulated to generate information. Information then leads to knowledge through interpersonal testing, validation and codifying. Another model is provided by Bellinger (1997) who suggests that understanding relations or association, understanding patterns and understanding principles are associated with the levels of information, knowledge and wisdom respectively. In his model, "…the sequence data → information → knowledge → wisdom represents an emergent continuum […] Everything is relative, and one can have partial understanding of the relations that represent information, partial understanding of the patterns that represent knowledge, and partial understanding of the principles which are the foundation of wisdom. As the partial understanding becomes more complete, one moves along the continuum toward the next phase."

As we can see from this brief review, then, there is no consensus on either the detailed definitions of data, information and knowledge, nor of their relationships. Definitions of 'data' range from context-free 'facts' to the intelligently-structured material found in databases, 'information' from Shannon's meaning-free signal to a contextualised and processed method of informing a user, and 'knowledge' from a further-refined variety of information to something that is attainable only through personalised experience.

When we look for commonalities in these definitions though, we find general agreement on two points. The first is that there is a process by which *something* is being transformed into *something else* that is more useful, either through a physical process of calculation or through a personal act of internalisation or contextualisation. A further point of agreement is that this processing is cyclic, with outputs becoming inputs to another process: this leads to the definitions of data-information-knowledge being fluid, shifting according to the perspective from which they are viewed.

However, the fundamental problem with this consensus is that it has at its heart a circular definition. Data, information and knowledge are each defined only in the context of their relationships with the other two, and it is impossible to separate the terms from one other. It is not actually possible to define at any point of observation of the business process whether we are looking at data, information or knowledge: if it is possible to view the same item of 'information' simultaneously as the expression of a value-added fact, as raw material input to a process, or as the basis for reasoned action, unless we have a frame of reference outside of the three definitions we cannot tell them apart. (How then, can we hope to be able to measure?) It is equally evident that the metaphor of transformation is flawed: processing does not create a new *type* of thing within the system, but rather shows an increase in the value of the material in return for the effort expended within a particular situation. All in all, the model begins to look rather like a Russian doll climbing an Escher staircase.

So what are we left with? What we can salvage from the definitions above is an agreement that there is a process of change in which an input is matched by a relatively valuable output – it is value-added in some way. Further, we can assess whether this processing has been successful or not in terms of the desired outcome of the business process (for example, we have reports generated from transactions, or graphs from statistics). Thus the extent to which the 'raw material' has been value-added can generally be assessed, at least intuitively, by a consideration of the *complexity* of the material at any stage. This is what we consider next.

NOETICA: A REVISED PERSPECTIVE

We begin from the one unquestionable fact that we have 'something' that is processed or acted upon to increase its value to the organisation. What we require is a term to refer to this 'something' that is (as far as possible) free from the intellectual baggage associated with the terms data, information and knowledge. We propose that since the material under consideration belongs to the realm of the intellect, the term *res noetica* (literally 'mental stuff') is appropriate. Although the term *res noetica* properly applies to anything that is a product of the intellect, in this paper we shall use 'noetica' to refer to all such materials as form the basis for computation, whether in digital form or as real world documents, procedures and practices.

We can now re-examine exactly what is involved in the process of adding value to this noetica and what form the concomitant complexification can take. Although, by definition, something is added to noetica when it is 'value-added', the term indicates our subjective appreciation of the process, not what actually happens. To examine what has happened we must look more closely at the changes in the noetica itself.

Value-adding typically takes the form of summarising (through tabulation or graphing), transforming (through selection, formatting or structuring), or adding context (through timeliness, relevance or incorporation into a bigger picture). Each of these activities involves an input into the system of skill, time and resources, which we may summarise as *effort*. The value-adding activities are intentional and the noetica itself is a product of intention, so the end result of each business process is an increase in the net intentionality embodied in the noetica. This increase in intentionality, a function of the state of the initial noetica and the input of effort, may be observed as an increase in the *order* in the noetica (in its sense as the opposite of chaos). Thus, we have a direct correlation between the value-adding process and a resulting increase in order in the noetica.

Let us consider what effect the addition of order has on the noetica. Starting with *simples* (noetica as it appears at the start of the process) and ending up with *complexes* (noetica made up of many simples) we can perceive three types of change:

- A process of *aggregation* forming new *composite* structures. Here the original simples are still visible, but are present as part of a new encapsulating structure. (The analogy here is from geology, where the rock type 'conglomerate' is made up of fragments of pre-existing rocks cemented in a matrix.) An example of an aggregation complex would be constructing a web site out of source documents, or making a master book document out of a set of chapter documents.

- A process of *transformation* forming new *compound* structures. Here the simples are no longer visible in the complex (although they still exist elsewhere in the system) - what has been created are new documents that derive from the simples. An example would be making a normalised set of tables from a number of spreadsheets, or a summary report out of transaction records.

- A process of *interrelation* leading to an increase in *contextualisation*. Here the simples are still present, but there are now connections made between them, and it is these connections and the networks that arise from them that are of value to the organisation. For example, we can run cluster analysis on a large number of documents to find commonality of referencing between them and infer chains of influence, but the original documents remain unchanged.

As the noetica builds up in ways normally associated with business computer processing, we see that rather than a postulated hierarchy, the result is a newly enriched noetica that is itself suitable for further processing. Any process, be it focussing on aggregation, transformation, or interrelation (or any combination of them), may take as input the output structures of any other process: simples become complexes by the action of enrichment, but will themselves always be potential simples for further enrichment.

This ultimately means that we will always have the possibility of further enrichment (given skills, time and resources) as long as we have a living organisation. So we have a process whereby there is a recursive action that adds to, rather than diminishes, the complexity (i.e. one where the limit set is infinite, rather than zero). This we term *allorecursion*, as it is the outwardly curving equivalent of the inward spiral that is conventional recursion. To illustrate the point, consider the archetypical spiral, the nautilus shell: on ex it might seem to be spiralling inwards, but the shell grows allorecursively as long as the individual nautilus has life.

Organisations interact at a noetic level at the same time as the standard business processes occur – the flow of communications and exchange documents, as well as the results of join ventures between them, ensures that this happens continuously. The organisations interact as noetic entities, and the same term can be applied to individual agents (people or organisational sub-units) or repositories of noetica such as libraries or document stores. There is a noetic universe comprising the sum total of all such material, and it is this unification of the entities within this universe that lies behind the inspiration of such visionaries as Bush, Mooers, Nelson and Berners-Lee.

Granularity, shape and scope

We have identified three different types of complexity that result from the order imposed by value-adding. We shall refer to the complexification due to aggregation as *granularity*, that due to transformation into new compound structures as *shape*, and that due to contextualisation through interrelation as *scope*. Let us now examine the dimensions of shape, granularity and scope more closely.

We find that an increase in *shape* through an act of will determines the set relatedness of the various items. The order that arises in the shape dimension resides in formal propositional structures, and with an increase in order, we can see a increase in the number of tables, indexes, data stores, views, and stored queries. And we see the simple structures (fields, tables) leading to low-order compound structures (databases, connections, stored multi-table queries) and then up to higher compound structures (data warehouses, data marts, ROLAP), acquiring increased shape allorecursively.

We find that an increase in *granularity* alters the way in which the noetica is perceived by the user, with new horizons becoming possible as new levels of complexity are reached. The order that arises is based in discrete structures, and unlike the allorecursion inherent in the shape dimension, we find that lower levels are occluded by the higher. Here, an increase in order sees bits and bytes in streams and on disks aggregate to form discrete composite noetica (for example documents, code or ADTs), which in turn become organised into higher level abstract structures such as directories, file systems, playlists, and so forth. At a higher level still, we are forced to use statistics (disk usage, hit rate on a site, the age of files, or the location of systems) to comprehend the aggregated noetica.

Lastly, we find that an increase in the number of potential connections leads to an increase in the *scope* of the noetica. The structures of the scope dimension are to be found in the organising principles that permeate the noetica, and so these structures are often much harder to isolate than those of granularity or shape. This is partly because scope manifests itself in the naming of fields, files and computers (since it is impossible to have unnamed aggregates or shapes) but also because scope is very bound up in usages and rules, and it is in the transient interaction with the user at particular points of the business processing life cycle that much of the scope is to be found. The results of this interaction are often an increased skill-level in the user, or a set of procedures in a written manual, and it may be from here rather than the digital domain that they will interact with the system the next time they are needed. We see scope embodied in structures such as classification schemes, procedure manuals, Petri nets, topic maps and ontologies.

Although described separately in the simple examples above, it is important to note that complexification cannot occur as a single type only. In each of the examples, we see complexification in all three dimensions: in constructing the web page there is an increase in contextualisation as well as aggregation; in the cluster analysis we also have a new compound structure as well as an increase in contextualisation; while the use of a thesaurus in a media archive will add a new shape to the system, and bind metadata values into each of the archived artefacts. Thus a typical business process will involve complexification along all dimensions simultaneously, but one or other is likely to predominate according to the current focus of interest.

By putting these structures in place (as file systems, backups, tables, indexes, reports, queries, views or ontologies) we ensure that the noetica can remain at a stable level of complexity, but it is a dynamic stability that we must maintain by actively maintaining the integrity of the structures.

The Noetic Prism

We can now summarise our revised perspective:

- We have used the term *noetica* to describe collectively all of the materials of computation (digital and non-digital), and argue that there is only one set of noetica, which is processed in different ways, and which acquires order through the imposition of three different principles.
- These three separate but interrelated principles of *granularity*, *shape* and *scope* each inform the noetica at any one time. The nature of any point in the noetica is characterised by the extent to which granularity, shape or scope is informing the decision-making or processing.
- The result of value-adding to the noetica is an increase in aggregate, compound and contextualised structures, measured as *complexity* of the noetica.

Using these principles, we can map the noetica in a 3-dimensional space framed by a triangular prism, which we call the *noetic prism*. The prism permits a *four*-dimensional co-ordinate vector space: the vertical axis represents complexity, which as we have seen is a measure of the intentionality stored in the noetica (as a function of time, effort and skill), and the three vertices represent the three dimensions of noetica – granularity, shape and scope (Figure 1).

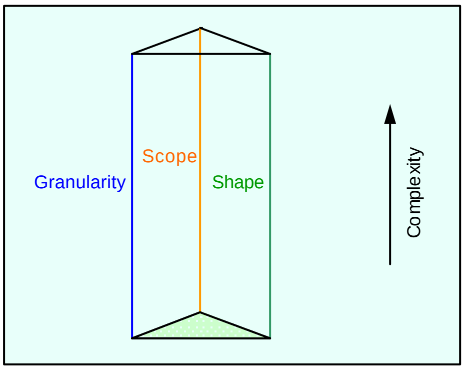

Figure 1. The Noetic Prism showing vertices of granularity, shape and scope.

Position in the vector space is given by the vector sum of the *importance* of each of the three dimensions. Importance is a vector of *extent* (displacement from a vertex) and *complexity* (displacement from the base of the prism). Each point $\overline{\mathrm{N}}$ in the noetic prism is thus determined by a 6-part value of

$$\overline{\mathrm{N}} = (e_{\mathrm{G}}, k_{\mathrm{G}}, e_{S}, k_{\mathrm{S}}, e_{C}, k_{\mathrm{C}})$$

where: e is a measure of the *extent* to which the vertex is significant at that point,
k is a measure of *complexity* and:

e_{G}	extent of granularity	k_{G}	complexity of granularity
e_{S}	extent of shape	k_{S}	complexity of shape
e_{C}	extent of scope	k_{C}	complexity of scope

As this is a vector, it reduces to a 4-part value of

$$\overline{\mathrm{N}} = (e_{\mathrm{G}}, e_{S}, e_{C}, k_{\mathrm{N}})$$

where k_{N} is the net value for complexity.

The space enclosed by the prism is fractal in nature and can be used to represent the noetic resource for a country, an organisation or an individual. It may also be used to represent a single item such as a database, spreadsheet, or mailing list. In each case the noetic resource has its own measure of complexity, scope, shape, and granularity, and the effect of adding a new body of noetica to an existing body can be measured as a vector addition:

$$\overline{\mathrm{N}}_1 + \overline{\mathrm{N}}_2 = \overline{\mathrm{N}}_{\mathrm{R}}$$

$$(e_{\mathrm{G1}}, e_{S1}, e_{C1}, k_{\mathrm{N1}}) + (e_{\mathrm{G2}}, e_{S2}, e_{C2}, k_{\mathrm{N2}}) = (e_{\mathrm{GR}}, e_{SR}, e_{CR}, k_{\mathrm{NR}})$$

The noetic prism permits us to model the noetica in several interesting ways, which we shall now examine. To begin with, any noetic state can be delineated in terms of the relative proportions of its noetica along the vertices of granularity, shape and scope: the impossibility of having extent along one vertex without extent along the other two is evident, since any point in the prism space will always have three vectors.

The prism shows how a ‘hierarchy’ of any sort is an inappropriate model for the disposition of the noetica, as the appearance of the noetica to any process at different points (in time or space) may be seen to be one of focus rather than separation. We can see how the emphasis of operations on the noetica shifts between the vertices of the prism with each consecutive analysis and operation: we cannot look at the noetica in all three dimensions at the same time, because we are always looking with speculative instruments that work primarily in one of the dimensions at the expense of the others.

In addition, greater complexity does not preclude further processing: this demonstrates the allorecursive cycle of old complex structures becoming new simple structures with increase in complexity along the prism vertices.

Finally, the use of a vector space to plot coordinates in the prism enables us (given the correct set of metrics) to employ vector mathematics to establish the effect of a certain effort (resource, skills and time) on any body of noetica. Thus when a process is carried out on the noetica, the coordinates of progress can be mapped, and the model can be used as a management tool to analyse and elucidate standard IT problems, as we shall consider later.

DATA, INFORMATION, KNOWLEDGE REVISITED

Our new model – the noetic prism – has as its vertices the dimensions of granularity, shape and scope. While we could continue to use these terms to characterise noetica, there are advantages in relating our model back to the terms information, data and knowledge. These terms are in common usage and unlikely to disappear: moreover, despite their often conflicting definitions we can find intuitive similarities between them and our model such that only a small shift in perspective is required to continue using them in the context of the noetic prism.

We propose the following revised definitions:

- *Information* as the term for noetica viewed along the *granularity* dimension
- *Data* as the term for noetica viewed along the *shape* dimension.
- *Knowledge* as the term for noetica viewed along the *scope* dimension.

We choose these definitions as the closest match between the existing terms and our model:

Data is noetica that has *shape*, as we are familiar with from databases and other data structures, where the structuring is allorecursive. Data can provide set information, can show the locus of interaction between the various sought results; it can represent the world in abstract data types such as queues, graphs or trees.

Knowledge maps readily onto the *scope* dimension, where reference to external standards, authorities, and communities of practice are paramount. Knowledge provides names, contexts, histories and above all purpose.

The mapping of *information* to *granularity* is perhaps less intuitive, but we consider information (as conventionally portrayed as an analysis tool) as inherently particulate, partaking of time and space and perspective. The occluding nature of information, whereby levels of abstraction determine level of focus, also supports this mapping. Information can check, verify, summarise, provide streams, give yes/no answers: it can locate at points in space and time.

The traditional hierarchy can now be seen as only one of the many possible noetica-enriching paths, although it is the one for which there exist many standard business processes. The revised hierarchies of Miller (1999) and Tuomi (1999) can be seen as alternatives that bring the accumulated contextualised noetica (viewed along the 'knowledge' dimension) to bear on the analysis of noetica, preparatory to its enrichment by acquisition and structuring. Extensions to the hierarchy – whether upwards to include ‘wisdom’ and ‘intelligence’, or downwards to include ‘events’ – can be seen to be a function of the allorecursive/recursive nature of the noetic substrate.

We can also see how ‘transformation’ through processing between levels of the hierarchy is explained by the problem of shifting focus between the dimensions as mentioned above. As we examine the noetica with a particular business tool, the noetica will take on a dimensional focus to match the conceptual framework of the tool. We begin processes with structures of a given dimensional complexity in focus, but with shift in focus, certain of that complexity is no longer available to us – it is not lost, but merely temporarily out of sight. Thus, in moving between information towards knowledge or data we lose both position and granularity; in moving from data towards knowledge or information we release shape and propositional forms occurring with shape; and in moving from knowledge to either data or information we lose scope and vision. It is this shift in focus that appears to be transformation.

MANAGEMENT IMPLICATIONS

One reason why the traditional hierarchical view of data/information/knowledge has been so widely accepted is that it is perceived to have direct payoffs in terms of analysis, modelling and planning on the one hand, and explanation and prediction on the other. To be useful, any revision of this view must offer at least the same payoffs. We describe the practical uses of the noetic prism next.

The noetic prism provides us with a straightforward mechanism for showing the status and direction of an IT operation, by giving an unambiguous and intuitive representation of its data/shape, information/granularity and knowledge/scope. We can thus use the prism:

- To describe a particular situation or state of an organisation, by determining the occupation of the three-dimensional prism space, using relative or absolute measures of complexity along the dimensions of shape, granularity and scope.
- To describe a problem and show what is required of a solution. Here we can show the dynamic behaviour of businesses processes as movement between the vertices of the prism.
- To avoid the complications that arise from working with structures particular to one vertex with tools appropriate to another, by being aware of how the focus of activity moves in the lifetime of a project.
- To plot courses of action and allocate resources. The management of any business computing process needs an accurate picture not only of the current state and the target state, but also of the stages through which the project must move. By mapping out the vector movement in the noetic prism, a plan for navigating from one to the other may be formulated.

To illustrate these uses, we now present analyses of three standard problems in business IT from the perspective afforded by the noetic prism. That is not to say that we offer a solution for these problems, but rather use the noetic prism to clarify the operations involved, and show how in effect there are category errors involved in the traditional analyses.

Case 1 - Mergers

The case of merging IT infrastructure takes several forms, for example:

- The merger of the IT capital of organisations on takeover, merger or joint venture
- The decision to bulk acquire new noetica (in the form of purchased data, knowledge bases and rule sets, or new adherence to an international standard)
- The problem of the re-appraisal or auditing of data models or data processing trails within an existing IT infrastructure

All of these examples are actually instances of merging noetic resources. Let us begin with the first instance, the organisational transition. Conventional wisdom has it that "mergers and acquisitions often produce results that are substantially less than the sum of their parts" (Smalley Bowen, 2000). The prevailing view is that incompatibilities of systems are the main cause for problems – indeed, Smalley Bowen recounts how some mergers have been called off simply because of the IT incompatibilities. The cost of takeovers, and the risk associated with them has led some business theorists to speak of the "knowledge-based merger", where "acquirers must develop an information support infrastructure that leverages their shared technology" (Sirower & Nicholson, 1999).

What form does this incompatibility take, and how can it be resolved? The division is seen as taking two forms: one systematic (software, operating systems etc) and the other, greater, one of implementation (existing applications developed within the systematic framework). In the case of the acquisition of Eastern Enterprises by Keyspan Energy in late 1999, advance preparation of the IT merger enabled the economic merger to proceed simply by not working towards a total unification of the systems, but adapting brokering technologies (via web interfaces) to keep the billing systems running (Smalley Bowen, 2000).

When we use the noetic prism to map out the process, we can see that the problem of incompatibility of implementation is caused by the attempt to combine two higher level sets of noetica that have been established and built up from lower levels of noetica through the expenditure of effort. In the Keyspan example, the two sets of lower-order simple noetica – billing of clients, installation of pipes, description of services – are clearly identical. It is in the process of modelling and implementation of those lower-order simples and the history of their transformation into higher-level complexes that the problem resides.

We expect the problem also because the ordering inherent in the noetica (which represents the intention transferred to the noetica via the expenditure of organisational effort) may take on more than one form of structure, and these different structures cannot reasonably be expected to match. This is because the noetica is ordered as (e.g.) knowledge trees which we have no reason to expect to match in terms, databases we have no reason to expect to match in shape, and aggregations with varying degrees of accuracy and abstraction. So effectively instead of one merger, there are three to contend with.

To formally unite the two sets of noetica, the problem reduces to a vector addition. And if the two vectors have an incompatibility in their structure, they have an opposition in their direction of their magnitude, and so the resultant vector sum is less than either. By working towards a proper solution, attention has to be paid to the reverse engineering of the granular and scope dimensions of the noetica to a commonality, and this involves the sacrifice of shape. The resultant noetica then needs additional expenditure of effort to return the shape of one of the two component parts to a level where it can be usefully employed as a billing system once more (either in terms of one of the two systems, or as a new system that can accommodate all of the noetica that comprised the two original sets). This process is illustrated in Figure 2.

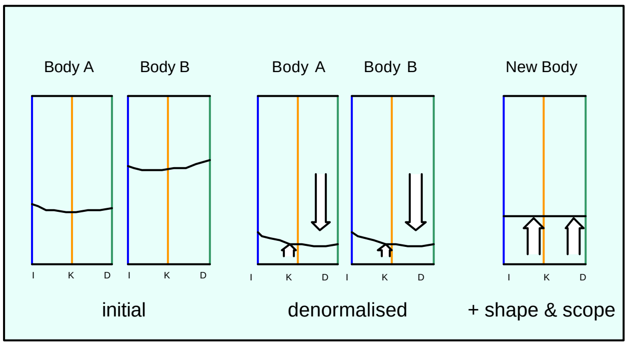

Figure 2. Merger of two disparate organisations.

While the examination of this problem from the perspective of the noetic prism does not offer any new solutions, it does help map out the reasons why the problem exists. It also maps out the stages that must be involved in a merger: firstly determining a commonality of purpose between the two systems and a common context for the work, followed by a process of denormalisation and disaggregation, and finally recreation of the new, combined set of noetica into a new higher order system. Smalley Bowen (2000) suggests that companies need to have IT-merger plans ready. It can be seen that an analysis based on the noetic prism may be of help here.

With the second instance (the bulk acquisition of noetica capital) the problem is similar; it is simply not as extreme. Here we face the prospect of (e.g.) buying mailing lists for merge/purging, buying up rent rolls, or acquiring thesauri. Generally, incoming noetica is acquired in a single block, it is smaller than the organisational noetica, and its focus is primarily on a single dimension. The solution is the same, however: recontextualising, denormalising, finding commonality and then adding order (mainly in the form of shape) to the new noetica.

Best practice business process involves adherence to recognised standards at the design stage, and this helps to minimise the problem. For example, using a common form for banking details, or buying a data system that is regularly used within an industry, or using an internationally-approved thesaurus, can mean that merge/purging with external sets of noetica is relatively-effort free. It was the availability of the MARC records in tape format that led to the prevalence of the OPAC throughout the world in the last decade. Here preparation of a library system to receive input in a standardised format meant that the cost of cataloguing was minimised, and uniform standards of excellence were established throughout the library system.

The third instance is the problem of the re-evaluation of an existing system, and the cost-benefit analysis that goes with any such re-evaluation. Here we may once more quote Calvin Mooers: Mooers' Law states that “an information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it” (Mooers, 1959). The problem again consists in the unravelling and re-ravelling of the order that has been put in place in analysis. It is for this reason that it is much easier to establish co-operation with standards at the beginning of an operation than half way through. This cost-benefit analysis is not always easy, as reputations are quite often welded to half-completed programmes of development, but the advantages of ‘buying effort’ (by paying for expertise, or by subscribing to a successful solution en toute, or by reshifting the focus to adhere to ISO standards) often outweigh the prospect of starting over.

Case 2 – Catalogue interaction: supply chain, supercatalogue and union catalogue

When it is necessary for two separate organisations each with a noetica capital to communicate for any purpose, there is generally a problem with co-ordination of data link-up and transfer. This is seen in three of the classic inter-organisational communication problems: the supply chain, the super-catalogue, and the union-catalogue.

The supply chain problem is typically one of getting just-in-time information from one organisation to the next in the chain in such a way that continuous operation will not be disrupted. “Here success implies that the system not only function as predicted, but more important, *fail gracefully* without disrupting other participants upstream and downstream” (Reddy, 2000). It is very much like an old bucket brigade, where the water will go from (e.g.) the stream or the well to the house on fire seamlessly as long as the rhythm is kept up. At its worse, though, it can be more like ‘Chinese whispers’, with each stage in the process compounding the errors of the previous one.

The super-catalogue is a system wherein a single request for information is to be met with by a search on multiple, disparate and diverse data sources. Here the problem is still timeliness, but the need for a single answer from asynchronous sources is the task that defies solution. The union catalogue takes this a step further, with wide-area catalogues for libraries, co-ordinated drug information systems, co-ordination of technical and scientific information from different agencies all requiring multiple data suppliers and users.

The problems of the supply chain and the super-catalogue are at first sight orthogonal to each other – in each one there are as many communications as there are organisations, but while the supply-chain has a single noetic translation process for each link, and each organisation has at most two links (input and output), the super-catalogue has the great burden placed on the single initiating system. The problem of the super-catalogue is made even greater when the individual organisations providing data are also the customers for it, as is the case with the union catalogue system. We shall deal with the additional problems of this scenario after describing the commonalities of the two styles of system.

The main task that faces the designer of either such system is that any disruption to the routine (of internal use) of the database-supplier will cause the supplier to not proceed with the system – here we see Mooers’ law in action once more. This is the so-called ‘zero-footprint’ requirement. Since the user of the data generally has a routine as well, it leaves one with the requirement for an efficient data brokering system. The actual IT problem (as opposed to the social, economic and political problems) soon becomes one that is similar to the problems described in the Mergers case. But in addition to having the initial bulk-adding and interpretation problems, we now face the problem of multiple small consecutive data accesses from disparate noetic entities. The problems are met with in different ways, but generally it is by an initial analysis and establishment of a brokerage system, to be followed by use of that brokering system on a needs basis. This can happen in one of two ways.

One conventional answer involves low-level brokering implementations such as CORBA. This solution has as its heart of an informationistic mechanism in the form of the abstracted CORBA mechanism. CORBA acts as a brokering agent between non-structurally redefined sources. This system suffers from the problem of concurrent loss as the contents of a data store get converted to aggregate noetica for transfer, and then must be converted back again at the other end. This is always going to be inefficient, but has the benefit of much greater flexibility than the alternative.

The second solution is the use of a data-language infrastructure, generally SQL92 in such systems as ODBC or JDBC. These solutions have a database system at their heart, and so require the loss of shape from any noetica that is not stored in tabular form.

Either way there is a loss of order, which has to be built up from the lower-order noetica that is transferred. The flow of noetica from the complex to the simple (which is transferred) back to complex once more zigzags back and forth between the point-defined information on the one hand and structurally-defined propositional (data) form on the other. Recalling that shape is gained at the cost of precision and topological information, we see how the act of categorising naturally occurring aggregation involves blunting it. This is illustrated in Figure 3. Here we see a loss of scope and shape, though not granularity, in the prism representing the stream of noetica as it leaves the source (left) to be accepted by a mediating device such as CORBA or ODBC. There is a greater loss of scope than shape and a decrease in value on the knowledge vertex (centre). When the noetica is reconstituted in the recipient (right) it regains shape and scope, but levels of knowledge, data and information are all lower than they were originally. This sequence is repeated along all the links in the chain.

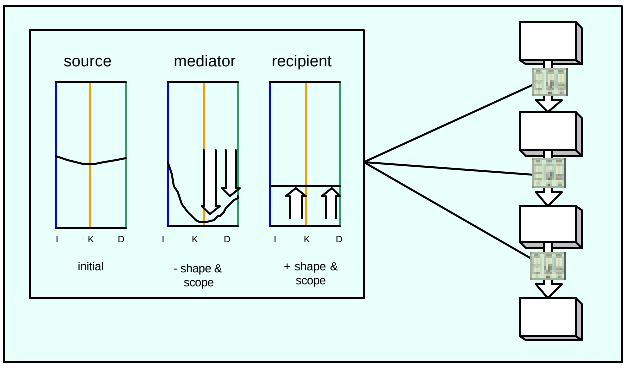

Figure 3. Supply Chain.

The standard view of the supply-chain and super-catalogue is a controlling perspective that offers an artificial point of view, which covers all of the resources embodied in the system. The ultimate reason we have these systems is that we need knowledge, and that knowledge is derived from the distributed points of information, from the stored rules and represented tacit assumptions, and from the gathered data sources which can be made by brokering various information points. At the same time we can see that the point source of all of information is the quantitative or qualitative data concerning events, stock levels, processes and outcome of those processes. In the standard abstracted view, we see received points of data brokered through a mediating data portal (such as CORBA, ODBC or UDA) and delivering the conclusions through stored queries that can advise the consultant manager.

When we plot the process involved in moving from the actual distributed points of recording up to the summary judgments that are required by management we can see how the different solutions of the supply chain problem actually follow different paths from the simple to the complex. These different paths match the different technologies and different methodology in the different ontologies that the various theorists and manufacturers have brought with them.

An additional problem that occurs with the union catalogue is that of ownership of the new system, and of the cost of creating it. Again referring to Mooers' law, there must be a perceived benefit if it is to be overcome. And if (as is often the case) there are rival sources for the starting point from several partially completed schemes, then the organisations that accept the abandonment of their scheme will suffer a loss of noetica capital as well as a loss of face. Dahn (1999) shows these problems clearly when he lists several alternative strategies for trying to establish a Whole-World union catalogue, including establishment, cataloguing and distribution challenges. Whichever way the problem is solved (i.e. by an information or data model) the stumbling block remains the impossibility of passing both the shape and the granularity at the same time.

Case 3 - Content-based retrieval

Content-based retrieval (CBR), whether by image, sound, or text is a fascinating example of the co-operation of the three vertices of the noetic prism. In CBR a match is sought between a standard indicator set and the media artefact that matches the search condition. In a 'conventional' database query, a match to a set-relational query will be retrieved by (e.g.) a drill down a b-tree, and the process of comparison is exact and immediate: what is returned is a set of pointers to zero or more record locations, and hence the record(s) displayed. The computer doesn't have to be told how to make the match. With a CBR-driven search, on the other hand, the process is analogous to training – initially the meaning is blank, and gradually the system is taught what certain patterns amount to within the context of a particular content domain. The correspondence of those patterns to features that can be described meaningfully is there only as the contingent product of having been confirmed.

In the terminology of Eakins and Graham (1999) we match *primitive* features such as colour or shape to correlate with *logical* features such as the identity of objects shown in the expectation that we will find *abstract* attributes such as the significance of the scenes depicted – and as Eakins and Graham put it "While CBIR systems currently operate effectively only at the lowest of these levels, most users demand higher levels of retrieval."

To make the leap between the logical and abstract features, standard references for (e.g.) distribution of colour and texture must be established, and then a process of training for that standard reference must occur. The standards are therefore established by the training to create a correlation between a principle of knowledge and a statistically determinable set of results. Thus noetic aggregations of shape and location are stored to match with contextualising sets that give semantic labels to features. As long as the system is well trained, the result can be acceptable within a given (trained) domain.

However, when the system is used to mediate between the user and the database (which is a structural, propositional representation of the subject matter) one ends up with the process of going from information → data (statistically likely) and data → knowledge (where each data point may cover more than one semantic label). But to assume that this gives more than a contingent correlation (and so only a likelihood of matching) is to make a category error.

Figure 4 illustrates this process. In a conventional query (left), the knowledge need is translated directly via a query broker to the database engine and then back again, and the levels of data and knowledge in the prism do not change. Content-based retrieval (right) begins with a stage of training to a reference set between the knowledge (concept) and information (media) vertices, and the second stage maps that trained occurrence with a record. These records are then queried using a conventional table-to-query process, and the retrieved media compared with the original concept. This inevitably involves a loss of scope due to the polysemous nature of media. The decrease in the knowledge vertex of the prism from its original level illustrates the overall inefficiency of the process.

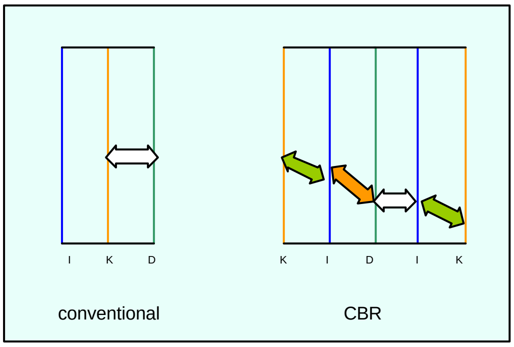

Figure 4. Conventional querying and content-based retrieval

CONCLUSION

We believe that our model of the noetic prism offers a useful new perspective on the old problem of definition (and consequent measurement and management) of the data-information-knowledge complex. By abandoning the hierarchical model of process and transformation, we are free to view the intellectual resources of an organisation and their use in terms of a focus on three different dimensions of complexity, seen as vertices of a triangular prism. This revised perspective allows a fresh analysis of many of the common problems in the management of intellectual resources, and we believe the noetic prism can become a valuable practical tool.

We note briefly that measurement and quantification of intellectual assets are seen as desirable as an indicator of the status of an organisation, or even a nation (Malhotra, 2000), and various measures of assessing data, information or knowledge assets have been proposed in the literature (Bontis, 2001). However, finding suitable metrics has always been problematical, not least because of the intangible nature of knowledge assets (e.g. Sveiby, 1998). Although a detailed treatment of the quantification of our model is beyond the scope of this paper, we believe that it will be possible to quantify complexification in the dimensions of shape, granularity and scope, and suggest that suitable metrics may be found in the literature relating to software engineering, quantitative ecology and dynamic systems, and group psychometrics.

REFERENCES

Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: conceptual foundations and research issues. *MIS Quarterly, 25*(1), 107-136.

Alberthal, L. (1995). Remarks to the Financial Executives Institute, October 23, 1995, Dallas, TX. quoted.in Bellinger (1997; op. cit).

Bellinger, G. (1997). *Knowledge Management — Emerging Perspectives*. Retrieved from: http://www.outsights.com/systems/kmgmt/kmgmt.htm.

Bontis, N. (2001). Assessing Knowledge Assets: A review of the models used to measure intellectual capital. *International Journal of Management Reviews, 3*(1), 41-60.

Clancey, W. J. (1997). *Situated Cognition: On Human Knowledge and Computer Representations (Learning in Doing)*: Cambridge University Press.

Clarke, R. (1999). *Fundamentals of 'Information Systems'*. Retrieved from: http://www.anu.edu.au/people/Roger.Clarke/SOS/ISFundas.html.

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM, 13*(6), 377-387.

Dahn, M. (1999). *Earth's Largest Library: One Librarian's Plan of Action*. *Searcher, 7*(7).

Date, C. J. (1991). *An Introduction to Database Systems*. (Fifth ed.). (Vol. 1): Addison-Wesley.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: how organisations manage what they know*. Boston: Harvard Business School Press.

Eakins, J., & Graham, M. (1999). *Content-based Image Retrieval* (39): JISC Technology Applications Programme. Retrieved from http://www.jtap.ac.uk/reports/htm/jtap-039.html.

Earl, M. J. (1994). Knowledge as strategy: reflections on Skandia International and Shorko Films. In C. Ciborra & T. Jelassi (Eds.), *Strategic Information Systems: A European Perspective* (pp. 53-69). Chichester: John Wiley & Sons Ltd.

Elmasri, R. A., & Navathe, S. B. (1999). *Fundamentals of Database Systems*. (3rd ed.): Addison-Wesley Publishing.

Hutchinson, S. E., & Sawyer, S. C. (2000). *Computers, communication and information: A user's introduction*. (7th ed.): McGraw-Hill Higher Education.

Malhotra, Y. (2000). Knowledge Assets in the Global Economy: Assessment of National Intellectual Capital. *Journal of Global Information Management, 8*(3), 5-15.

Masuda, Y. (1981). *The information society as post-industrial society*: WFS. Tokyo.

Miller, F. J. (1999). *I=0: Information has no intrinsic meaning*. Retrieved from: http://www.sveiby.com.au/miller99.htm.

Mooers, C. N. (1950). The theory of digital handling of non-numerical information and its implications to machine economics. *Zator Technical Bulletin, 48*.

Mooers, C. N. (1959). Mooers' Law; or why some retrieval systems are used and others are not. *Zator Technical Bulletin, 136*.

Newell, A. (1982). The knowledge level. *Artificial Intelligence, 18*(1), 87-127.

Newing, R. (2000). *Business intelligence: A fundamental pillar of knowledge - and of wisdom.* FT.com. Retrieved from: http://specials.ft.com/ln/ftsurveys/industry/sc70aa.htm.

North, K. (1999). *Excellence in database technology*. Retrieved from: http://ourworld.compuserve.com/homepages/Ken_North/db_hall.htm

Polanyi, M. (1966). *The Tacit Dimension*. Garden City, N.Y.: Doubleday.

Pór, G. (2000). *Knowledge -> Intelligence -> Wisdom: Essential Value Chain of the New Economy*, [Keynote address delivered at the Consultation Meeting on the Future of Organisations and Knowledge Management of the European Commission's Directorate-General Information Society Technologies, Brussels, May 23-24, 2000]. Retrieved from: http://www.co-i-l.com/coil/knowledge-garden/kd/kiwkeynotes.shtml.

Reddy, R. (2000). Building the Unbreakable Chain. *Intelligent Enterprise, 3*(3). Retrieved from: http://www.intelligententerprise.com/000209/feat3.shtml

Roszak, T. (1986). *The Cult of Information: The Folklore of Computers and the True Art of Thinking*: Lutterworth Press.

Sammet, J. E. (1969). *Programming Languages: History and Fundamentals*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423 and 623-656.

Simon, H. (1969, 1995). *Science of the Artificial.* (3 ed.).

Sirower, M. L., & Nicholson, G. (1999). Knowledge-Based Acquisitions. *Banking Strategies Magazine, LXXV*(V). Retrieved from: http://www.bai.org/bankingstrategies/1999-sep-oct/Articles/M_A_Forum/

Smalley Bowen, T. (2000). Plugging IT into the merger equation. *Info World*. April 21, 2000. Retrieved from: http://www.infoworld.com/articles/hn/xml/00/04/24/000424hnmerge.xml

Sveiby, K.-E. (1998). *Measuring Intangibles and Intellectual Capital - An Emerging First Standard*. Retrieved from: http://www.sveiby.com.au/EmergingStandard.html.

Sveiby, K. E. (1997). *The new organisational wealth: managing and measuring knowledge-based assets*. San Francisco: Berrett-Koehler Publishers Inc.

Tuomi, I. (1999). *Data is more than knowledge: Implications for the reversed knowledge hierarchy for knowledge management and organisational memory*. Paper presented at the 32nd Hawaii International Conference on System Sciences, Hawaii.

Vannevar Bush. The Atlantic Monthly; July, 1945; As We May Think; Volume 176, No. 1; pages 101-108.

Weaver, W., & Shannon, C. E. (1949). *The mathematical theory of communication*. Urbana, Illinois: University of Illinois Press.

Weiner, N. (1948). *Cybernetics*: MIT Technology Press.